Proposed Pilot Study Design

Report from the Pilot Study Design Team and Technical Work Group

This proposed design was agreed upon by the FACDQ at the July 13-14, 2006 public meting at Arlington, Va. It informed the contract scope of work sent to potential bidders. The actual pilot will be conducted under an EPA contract.

I. Introduction/Background

At the March 29-30, 2006 Federal Advisory Committee on Detection and Quantitation Approaches and Uses in Clean Water Act Programs (committee) meeting, the committee assigned a Pilot Design Team (Team) – comprised of one Technical Work Group member from each caucus (See Attachment A for representatives) – to develop a pilot study design. The committee agreed, by consensus, to task the Technical Work Group and this "Study Design Team" with scoping the details of the pilot study. The committee also agreed by consensus to proceed with pilot testing the following five analytical methods: 200.7 (metals), 300.0 (ions), 625 (SOCs), 608 (PCBs, pesticides) and 335.4 (cyanide).

The committee instructed the Team to use two committee documents – "Revised Draft - What do we need a procedure to do" and the "Draft Pilot Study Purpose and Objectives" (Attachment B) – in designing the pilot.

It was understood that, while EPA is providing substantial funding for this effort, funding is limited, and trade-offs would be necessary. Additionally, the 6-month or less time period available to conduct a pilot study and required analyses posed limitations.

The Team met in Denver, Colorado on April 27th and 28th to develop a draft design. This document is the result of that effort.

II. Basic Design

A. Options for Pilot Testing Procedures

A key aspect to the pilot study is a need for it to produce data and information that will be useful for the committee in developing recommendations, and to have the pilot do this in a very limited timeframe. The Team discussed whether or not a pilot can be completed within a time frame that allows for data production and analysis to the level necessary for usable results. The Team considered four time period sequences for the pilot study (Attachment C). In summary these options are:

- Option A-1: The committee will make decisions around uses and conduct a pilot test, which will both lead to final committee recommendations. After the committee charter expires in May 2007, a follow-up pilot test will be conducted to confirm the committee's recommendations related to a

procedure(s) and the committee's recommendations will be used for rulemaking.

- Option A-2: The committee will make decisions around uses and conduct a pilot test, which will both lead to final committee recommendations. If delays occur and laboratory or data analyses for the pilot are not completed within the 6-month time window, a parallel approach will be taken. The committee will move forward with recommendations while pilot efforts continue. Some pilot results may be available late in the committee's development of recommendations or after the committee's recommendations are complete. A confirming pilot would still be anticipated after the committee makes their recommendations and the first pilot is complete. The rulemaking process would follow a confirming pilot.
- Option B: The committee will make decisions around uses that will go directly into the committee's recommendations. A confirming pilot test will be conducted after the committee charter expires in May 2007, which will be followed by rulemaking.
- Option C: The committee will make decisions around uses and conduct a "mini" pilot using only volunteer labs, and then make final committee recommendations. After the committee charter expires in May 2007, a confirming pilot will be conducted, which will be followed by rulemaking.

The Team recommends conducting the pilot now because this will provide the committee with the information it needs for decision-making. The Team agreed that the schedule, while tight, is feasible. If issues arise, the decision to conduct a full pilot study now should be revisited, and could include a consideration of extending the committee charter.

The Technical Work Group (TWG) plans to produce a separate document for the committee that will detail what this pilot study will not do. The goal of that document is to help build the framework for a post-committee confirmation pilot study.

B. Summary of the Basic Elements of the Pilot Study

The following provides a summary of the basic elements of the two parts of the pilot study: the regression design and the single-laboratory design (see Attachment D). All labs participating in the pilot will run samples under both designs with an exception for the Aroclors in EPA method 608, which is explained in the next paragraph. It is not necessary that all labs run all analytical methods.

Differences between single and interlab procedures are: (1) labs will prepare their spikes and calculate their detection and quantitation limits under the single lab design; (2) single-blind spikes will be sent to labs under the regression design; (3) the Team will calculate a detection and quantitation limit from the results of the analyses of the regression spikes using the ASTM IDE and IQE procedures; and

(4) labs that choose to only bid on the two target Aroclors (1016 and 1260) in method 608 will do so only under the single lab design. This exception for Aroclors is made to conserve resources and take advantage of existing Michigan Manufacturers Association (MMA) Aroclor data. Note first preference for Aroclor analysis will be given to labs that participated in the original MMA study. Various schematics are attached that may assist in visualizing the components of the design (Attachments E, F, and G):

Regression Design

- Minimum of 8 labs
    - Labs will be solicited for interest and must pre-qualify in order to bid.
    - Pre-qualified labs may bid on one or more methods.
    - The 8 qualified bidders that give the best value to EPA will be selected for each method.
        - Labs should be representative of population using method as much as possible (e.g., inclusion of small, medium and large labs).
        - Labs should submit bids based on a total price, rather than price per sample (total samples for single-lab pilot unknown at start of study).
- Five analytical methods
    - EPA Methods 200.7 (metals), 300.0 (nitrate, ions), 335.4 (cyanide), 608 (Pesticides) and 625 (organics).
    - Analytes listed in both 608 and 625 will be analyzed by 608 only.
    - Attachment I (Target Analytes for Pilot Study Design) describes the analytes the Team proposes to test.
- Historical blank data collected from labs
    - Analyte data generated during last 30 analytical batches or last 6 months, whichever yields the greater number of results from the instrument(s) used in the study.
    - Data generated on the same instrument will be used in the study.
    - Report blank data without any reporting limit censoring; may require labs to review/revise their historical data.
- A range of concentrations will be analyzed for each method
    - Estimate 12 concentrations, including a blank sample.
    - Exact details to be determined by Team based on lab proposals during the pre-qualification stage (each lab will review the LC-MRL procedure and state which spike levels they will use to perform the procedure; the Team will choose spike levels to reflect lab responses as much as possible).
    - Concentrations should approximate those needed to determine procedures, such as ACIL, IDE, IQE, LC-MRL.
- Ten replicates at each concentration by each lab for each method
    - Concentrations will be blind to labs.
    - A spiking lab or standards vendor will prepare and label each sample.

    - Samples will be based on the study spiking scheme approved by the Team.
- PCB Aroclors will not be evaluated for regression-based procedures using new data
    - Existing data with appropriate design is available from MMA PCB dataset.
    - Limits will be calculated and confirmed using the same approach that will be used to evaluate these limits with pilot study data to the extent possible using the MMA design.
    - While Aroclors are not included in the regression design, additional analyses of two Aroclors at three concentrations will be analyzed for use in confirmation of single-laboratory limits (see below).

Single Lab Design

- One single-laboratory procedure (Modified ACIL - Revision 5.1) will be evaluated by each laboratory independently
    - Each laboratory will choose initial spike level, prepare samples, analyze samples and determine the limits.
    - Both start-up (seven replicates) and ongoing limits (based on twenty replicates) will be calculated.
- Five analytical methods
    - EPA methods 200.7 (metals), 300.0 (nitrate, ions), 335.4 (cyanide), 608 (Pesticides) and 625 (organics).
        - Two Aroclors (1016, 1260) will also be analyzed using Method 608 (confirmation will be based on additional laboratory analyses at multiple concentrations by each laboratory).
        - MMA PCB data inappropriate to apply the single-laboratory procedure.
    - Attachment H (Target Analytes for Pilot Study Design) describes the analytes the Team proposes to test.

III. Elements of the Pilot Design

A. Lab Pre-qualification

In order for the pilot to be completed in a timely fashion, it is necessary to begin pre-qualification steps immediately. EPA can then select the labs who will be involved in the pilot study by early summer. For the lab selection process, the Team developed pre-qualification criteria. The Team discussed their desire to include a variety of labs in the pilot study. The Team agreed, however, that the extent to which full laboratory representation is achieved is contingent upon responses to a Request for Proposals, the bids received, and the pre-qualified laboratories. The list of criteria is shown below:

Lab Prequalification Requirements

#	Criteria	Reason
1	A statement on the # of analyte analyses the labs can perform using the selected method(s), and # typically performed per day	Assess lab's experience and confirm lab can realistically meet study schedules
2	Info about: make and model of the instrument(s) and detectors that will be used in the study, the current calibration period, cleanup and sample introduction procedures used,	May help explain any trends observed in data during study
3	A statement on how often the lab typically calibrates their proposed instrument(s)	Evaluate effect of instrument stability; possibly select a group that represents varying degrees of instrument stability
4	Copies of all method-specific lab QC data (blanks (uncensored for 200.7/335.4), calibration, spiked QC, etc) generated during last 6 batches or 30 days, whichever is greater. Must be on same instrument used in study	Verify that the lab is proficient with requirements of the method
5	Data from lab's most recent performance evaluation sample analyses, and/or certified reference materials	Determine how well lab analyzes QC samples of unknown concentration
6	List of staff who will participate in study, (include roles, responsibility, experience)	May help explain any trends in data collected during study
7	A series of potential spike levels that lab believes would be appropriate for use in determining LC-MRL and single lab procedure	Will be used by Team to select the study spiking scheme that will be used by all labs
8	Certification of the accuracy of their pre-qual package and their willingness to meet study schedules	Verify labs understand importance of study, design, and schedules and are willing to commit to meeting them.
9	Information about the lab's current raw data and actual methodology to arrive at a detection limit. This includes reporting any MDL, ML, L_C, L_D, and/or L_Q determinations by analyte, method, and instrument.	To understand the current practices used in each lab.

#	Criteria	Reason
10	The name(s) and nature of any multi-lab or round-robin studies in which the lab has participated in the last five years	To assist in determining the representativeness of the labs and to determine if past performance is satisfactory. It needs to be clear, however, that this is not a requirement that a lab will have participated in a past study

B. Procedures

The committee agreed to a list of candidate procedures for pilot testing. They indicated that it would be acceptable to consider modifications that a procedure developer wanted to propose. The following list includes those procedures that the committee included as candidates (See EPA web site at http://epa.gov/waterscience/methods/det/):

Committee Candidate Procedures

Detection	Quantitation
EPA MDL	EPA ML
ASTM[1] IDE	ASTM IQE
EPA Hubaux-Vos	EPA Lowest Concentration-Minimum Reporting Level
ACIL Proposed Procedures for Determining the Method Detection Limit and Minimum Level	ACIL Proposed Procedures for Determining the Method Detection Limit and Minimum Level
Consensus Group Proposed Procedures for Estimating the Critical Level and Quantitation Limit	Consensus Group Proposed Procedures for Estimating the Critical Level and Quantitation Limit
East Bay MUD Procedure for Determining a Detection Limit using Lab QC	

Because of resource and time constraints, the committee understood that the pilot study would be limited to a maximum of three detection and three quantitation procedures, and they requested that the TWG narrow the procedures to be tested to the understood number, for ultimate concurrence by the committee at their July 2006 meeting. The Team recommends the following procedures for testing as part of the pilot:

Approach	Detection Procedures	Quantitation Procedures

[1] American Society of Testing and Materials (ASTM)

Interlab Procedure – tested under regression design	ASTM IDE	ASTM IQE
Single Lab Procedure – tested under regression design	EPA Hubaux-Vos	EPA Lowest Concentration-Minimum Reporting Level
Single Lab Procedure – tested under single-lab design	ACIL Proposed Procedures for Determining the Method Detection Limit (ACIL) and Minimum Level	ACIL Proposed Procedures for Determining the Method Detection Limit and Minimum Level

Note (1): Although the Hubaux-Vos and LCMRL are single-lab procedures, they will be tested using the spikes developed for the IDE and IQE under the regression design.

Note (2): Detection or quantitation procedures that provide interlaboratory estimates for detection and quantitation differ from the other procedures in the table that provide intralaboratory estimates.

"Interlab" refers to calculating one limit from the results of experiments in several labs, with that limit calculated based on interlaboratory variability. An example is the IDE/IQE that will be tested in the regression design of the committee pilot study.

"Intralab" or single-laboratory refers to calculating a limit from the results of experiments conducted in one lab. An example of this is the ACIL procedure that will be tested in the single laboratory design of the pilot. Any set of single laboratory limits may be pooled to calculate a multi-lab limit, based on the mean and variability of the different single-lab limits.

In making their recommendation for choosing between the ACIL and Consensus Group procedures, the TWG compared each of the following attributes developed by the Team:

Attributes	**ACIL**	**Consensus Group**
Status	Complete but changes could still be made to meet other needs	New draft proposed but not completed
Statistical Factors	t=3.14 (n=7) t=2.54 (n=20)	k=6.101 (n=7) k=3.67 (n=20)
Test Frequency	Quarterly test	Every batch
Simplicity	Simpler	More complex

Attributes	ACIL	Consensus Group
Pilot Ease and Cost	Easier and less costly	More complex and costly
Spike Levels	Not defined	Defined

Technical Work Group Decision on ACIL or Consensus Group Procedures

The TWG evaluated the ACIL and Consensus Group procedures to decide which procedure would be tested in the pilot. One key point discussed during two sessions (May 10th and 12th) revolved around the following questions:

- Does the FACDQ want a procedure that labs would use to demonstrate they can achieve a specified data quality at their reporting level?

Or

- Does the FACDQ want a procedure that identifies the lowest concentration that a lab can achieve a specified data quality?

The ACIL procedure being considered focuses on a lab demonstrating what it can achieve while the Consensus Group procedure comes closer to identifying the lowest concentration.

Balanced with this discussion was a general sentiment, though not a consensus, that it would be important to test the simpler of the two procedures. A majority of the Team agreed the ACIL procedure met this criterion.

Within the points raised above, most of the TWG also preferred testing the ACIL provided that spike levels would be low enough, especially for the PCB analyses. Because the pilot design includes spikes at this lower range, it was understood that low levels would be addressed even though the ACIL does not specifically require this.

Two additional changes were raised that the TWG agreed by consensus to incorporate into the ACIL procedure:

1. For calculation of L_C and $L_{Q,}$ the mean of sample blanks analyzed will be set to zero if the means is less than zero.
2. The factor "t" will be changed to "k", and a multiplier of 2 will be used for L_Q. The multiplier used in the formula for L_C will be changed from a t-statistic to a tolerance limit multiplier k. This will change the resulting Lc from a prediction limit for a single result to a 99% tolerance limit for 99% of the population of method blanks. In addition, the requirement that L_Q be three (3) times L_C will be changed to a requirement that L_Q be two (2) times L_C.

With these changes, the TWG reached consensus that the revised ACIL procedure (Version 6.0) will be used in the pilot as the single lab procedure.

C. Measurement Quality Objectives

The committee set measurement quality objectives (MQOs) for the pilot test at its March 2006 meeting. The committee decided accuracy would be determined by looking at experimental data for each of the analytical methods. The Team agreed on several existing studies that will be reviewed with MQOs in mind. The following address the items to consider:

- Investigate target MQOs for accuracy using existing data
    - Department of Defense (DOD) Quality Systems data
    - Existing Lab Data from Severn Trent Lab data
- Analyze existing data for pilot study methods
    - 608
    - 625
    - 200.7
- Include mean and variability of recoveries for multiple laboratories
- Examine estimated MQO for each analyte
    - 95^{th} percentile of laboratory means
    - Mean + t*s (alpha for t = 0.05)

D. Laboratory Analysis Sequence

The following bullets list the parameters around which the spike samples will be tested by the labs for the Regression based pilot:

- Laboratory analysis sequence will be controlled using various options.
- Sequencing strategy is intended to ensure sample concentrations are blind.
- Comply with spirit of ASTM D-2777 requirements regarding randomization and avoidance of carryover.

E. Laboratory Analysis Timeframe

A tightly controlled sample laboratory analysis timeframe is necessary to ensure pilot study data can be analyzed and interpreted by the Team and the TWG for use by the committee in a timely manner.

- Lab schedules will be strictly controlled.
    - Samples must be analyzed over 4-5 weeks.
        - 45 calendar day maximum will be allotted (approx 15 day cushion for labs).
        - Labs must meet holding times.
    - Samples for the regression pilot will be prepared and distributed in weekly batches.
        - All batches must be analyzed at least 24 hours apart.
        - At least 2 batches per week must be analyzed on non-consecutive days.

    - Samples for the single lab pilot will be prepared by the laboratory.
        - Seven samples are analyzed at the labs projected L_Q initially, and then at least one sample on at least 20 separate days within the 45 calendar day maximum.
- Laboratory analysis anticipated from September 5 – October 20, 2006
    - Prior steps include:
        - Finalizing study design;
        - Development of lab Statement of Work (SOW) and instructions;
        - Lab pre-qualification, solicitation, and award;
        - Finalizing study spiking scheme;
        - Soliciting/awarding a spiking vendor; and
        - Vendor preparation of spiked study samples.

F. Lab QC

The Team agreed the labs need direction related to quality control. The following items list those elements of quality control that need focus:

- All laboratory analyses must be performed on a calibrated instrument.
- Labs will report if they recalibrate during study.
- Labs will follow the calibration requirements in the method.
- All method-specified lab blanks must be analyzed before each batch.
- Reflecting routine analysis, blanks should be as free from contamination as possible.
- Labs will follow only relevant method-specific lab quality control requirements.

G. Laboratory Analysis Requirements

In order to provide the most neutral of conditions for the pilot, the following items are included by the Team related to laboratory analysis requirements.

- Samples must be carried through all preparation and laboratory analysis steps as are typically used for wastewater samples, such as:
    - Digestion, extractions, and cleanups;
    - Instrument parameter set ups; and
    - Laboratory staff that conduct the study laboratory analyses be the same staff that routinely conduct laboratory analyses by that method.

H. Data Reporting

In order to evaluate the data effectively the Team identified the following data reporting requirements.

- Labs should not censor any results for which the instrument yields a numeric result. This means the laboratory should report even negative values or values less than the laboratory's current reporting limit.
- Labs should identify any qualitative identification criteria they use that differ from the criteria specified in the EPA method, e.g. criteria used to identify and quantify analytes using GC/MS.
- Labs will report run logs weekly to the prime contractor ("the contractor") to allow monitoring of study status, holding times, and laboratory analysis sequences.
- Labs will report summary level electronic data by week to the contractor, beginning 14 days after completing first week of laboratory analysis, and concurrently provide all supporting raw data. Raw data includes peak areas for calibration data and analytical runs.
- A standardized electronic format will be developed & provided to labs.
- Labs must use this standardized format to expedite data review, data distribution, and data analysis.
- Labs should retain raw data for a period of 5 years and provide it on request (and at additional cost negotiated as necessary).

I. Data Inspection

In order to provide the ability to quickly evaluate lab generated data the Team recommends certain data reporting provisions. The Team also desires to have some amount of the data reviewed. The following items address these concerns.

- The contractor will use standardized checklists to inspect summary level data to confirm that:
  - Results are correctly formatted (The contractor will work with labs to correct problems).
  - The lab analyzed each batch as directed using the appropriate method and within holding times.
  - Samples were analyzed on a properly calibrated instrument.
  - Method-specified blanks were analyzed with each sample and that potential effects of any contamination were reported.
- The contractor will randomly select and review 5% of raw data per lab.
- The contractor will compare raw data results to those reported, and for censored methods, such as EPA 608 and 625, check that all qualitative identification criteria have been met when results are reported.
- The contractor will confirm that method-specified calculations were properly performed.
- The contractor will confirm electronic data accurately reflect raw, hardcopy data.
- The contractor will conduct a more in-depth raw data review if warranted by findings.

- The contractor will ensure that clean-up procedures were used where required.
- The contractor will report data inspection findings to the TWG.

J. Data Distribution

The Team established that the data will be evaluated by the contractor and then by the TWG. This section also details who will generate the detection and quantitation limits.

- Electronic database will be distributed to the TWG after the contractor inspection/verification of lab submissions.
- To minimize version control problems, the contractor will distribute a complete database of all data submitted within study schedules.
- Participating labs calculate their detection and quantitation limits where possible (e.g. single lab procedures).
- The prime contractor calculates limits for pooled LCMRL, IQE, and IDE data, and checks lab's calculations for single-lab procedures.

K. Data Analysis

Data analysis is a key component of the pilot study that occurs after the data are generated. It is also important to note that there could be an unlimited number of data analysis processes that could be included. The Team wants to better define what data analysis will be done.

Because of limited resources and the timing constraints of evaluating the data from the pilot, the committee needs to make decisions about what kinds of data are going to be informative. The committee will need to make this decision before it starts to receive the data. The Technical Work Group is preparing a separate document that will "tee up" for the committee the data analysis issues that need direction.

L. Confirmation for Single/Multi/Inter Lab Procedures

A key element of the pilot will be the confirmation of the procedures. The Team has established the following criteria for confirmation.

- Limit calculation as discussed above will not require all data analyses at each spike level. Additional data analyses may be necessary to assess whether calculated limits achieve target MQO criteria. The following should be considered.
    - Percent RSD
    - Mean and/or Individual Recovery
    - False Positive Rate for uncensored methods by using blank data,

    - False Negative Rate (based on procedure's L_C and/or instrument signal)
- Data will be divided into sets for limit calculation (LC-MRL, Hubaux-Vos, IDE and IQE) and limit confirmation randomly or systematically, assuring that full temporal variability will be covered in both sets.
- For confirmation of L_C for uncensored methods, existing blank data may be used. L_C for censored methods, however, will not be confirmed because the definition and use of L_C for these methods is not clearly defined.
- It is unlikely that determined limits will match one of the spike levels exactly. Therefore, model fitting will be necessary. Models based on confirmation data will:
    - Estimate MQO values at determined limits.
    - Estimate concentration at which target MQO criteria are achieved.
    - Compare limits to MQO concentrations and each other where applicable.
- For the Single-Lab procedure, confirmation will be done using regression-design confirmation replicates analyzed by that lab.
    - Additional laboratory analyses for two Aroclors (1016/1260) by 608 will be needed for confirmation (five replicates each containing both Aroclors for three spike levels surrounding determined quantitation limit).

M. Pooling Data

The Team agreed that more specific criteria for pooling lab data to derive a single detection and quantitation limit from limits calculated by several laboratories were appropriate. This section summarizes EPA's approach for pooling detection or quantitation limits developed by several laboratories using single-laboratory detection or quantitation procedures, such as the ACIL, Consensus Group, or the EPA MDL procedures. The following single-laboratory approach may also be applied to procedures, such as the LC-MRL quantitation limit and the Hubaux-Vos detection limit calculations.

<u>Pooling Single Laboratory Data</u>
ASTM Standard D-2777 describes how to determine the precision and bias of an analytical method from analysis of pairs of samples spiked with several concentrations. This standard allows removal of results from a laboratory if those results are inconsistent with results from all laboratories in the study, followed by removal of outliers using statistical outlier procedures such as the Grubbs' test. In developing pooled detection limits from MDL studies, EPA follows the spirit of ASTM D2777 with respect to screening labs and data, but spiked pairs clearly are not used to calculate an MDL.

For this pilot, each laboratory's data would first be screened as a whole for conformance to both the detection or quantitation procedure, and the analytical method. Based on this screening, data from any laboratory that was not in control based on failing method QC, or did not run the detection or quantitation procedure properly, would be categorized as invalid and not included in a pooled detection or quantitation limit calculation. To accommodate concerns about outlier removal, we would omit this step in the primary analysis, but retain the option to include it, as appropriate, in the secondary analysis.

From the distribution of the single-laboratory detection and quantitation limits a mean and standard deviation would be calculated from these individual laboratory limits. The final detection or quantitation limit could be specified by taking this mean value plus a term, such as three standard deviations (i.e. the final limit would estimate the detection or quantitation limit that 99% of labs would achieve). If the single-laboratory detection or quantitation limits do not appear to follow a normal distribution, the pooled limit could be determined by taking the maximum of the individual laboratory detection or quantitation limits.

Interlab Procedures

The ASTM IDE and IQE detection and quantitation procedures will not pool data as such. The study will produce an interlab estimate of detection and quantitation limits from an analysis of the spike levels developed by the study designer.

N. Role of Existing Data (e.g., MMA PCB data)

This section includes consideration of existing data in providing additional information.

- Method 608 was used in the MMA PCB study; therefore Aroclors will not be included in regression-design study.
    - Multiple labs, spike levels, and replicates.
    - Laboratory Analyses conducted over six months
- MMA PCB study data can not be used for assessing Single-laboratory procedure, so additional laboratory analyses necessary.
- Each laboratory will analyze two Aroclors when performing single-laboratory procedure.
- Each laboratory will analyze five replicates at three concentrations for use in confirmation laboratory analyses.

O. Analyte Selection

This section describes some of the factors used to select the analytical methods chosen by the committee at the March 2006 meeting. Attachment H (Target Analytes for Pilot Study Design) describes the analytes the Team proposes to test.

- No Suitable Existing Data: The five EPA methods selected for the pilot study (with the exception of the Aroclor (PCB) data discussed in Section N) either

had no existing data, or that data was not at appropriate spike levels to assess limits of interest.
    - Method 200.7 – Rev. 4.4, 300.0, 335.4: only blank or LCS data available, no multiple spike-level data.
    - Method 625: no available existing data.
- Selected analytical methods are widely used, covered a range of chemical classes, include censored and uncensored methods.
- Selected methods diverse enough to attract a range of laboratories.

P. Priorities

While the Team felt comfortable with the pilot design decisions they made, they still recognized that a bidding process and unknowns could cause costs to increase. In anticipation of such a possibility, the Team prioritized what elements of the pilot should be dropped from the study. The following list details their priorities:

1. Drop EPA Method 300.0 (anions), and see if volunteer labs might run it.
2. Drop EPA Method 200.7 (metals) for single lab.
3. Drop EPA Method 200.7 (metals) altogether. Rely on the extensive amount of metals data provided by FDEP and other labs.

If the bidding process causes costs to decrease, and additional funds are therefore available, the Team recommends inclusion of additional laboratories in the pilot study.

Note: changing any other parameter, such as procedures to be tested, number of spike concentrations and replicates would require a rebid of the pilot design, and delay the start of the pilot.

IV. Resolution of the Trade-offs

The Team evaluated each of the elements of the pilot design with the available funding in developing a pilot study approach. The result of this evaluation was that conducting an ideal pilot was not possible when considering the substantial, but not unlimited, funding situation, so that trade-offs were needed. The Team established an iterative process for resolving these trade-offs. They considered what the committee needs a procedure to do as they worked through the details of the pilot design. The Team also kept in mind the limited time period available to conduct the pilot, including consideration of what would be practical to accomplish in the time period available. The result of the Team's discussion is a pilot approach that is anticipated to provide valuable information and data that can be used by the committee in developing their recommendations.

The elements that the design team focused on in developing a pilot approach included:

- Analytical methods and analytes to be tested.
- Costs for conducting specific analytical methods.
- Number of repetitive tests for each spike level.
- Number of spike concentration levels for each analytical method.
- Number of laboratories conducting tests.

Through an iterative approach the Team was able to reach agreement on a design. Attachment D is a summary of the worksheets that the Team used in their discussions. The following explains the Team's rationale and resolution of the trade-offs.

The draft study design submitted by the Regression Design subgroup included a minimum of nine paid laboratories analyzing seven replicates at 13 different concentrations. However, this design did not include any confirmation sample laboratory analyses (i.e., sample laboratory analyses used to test whether the calculated limits achieved their intended MQO criteria), and could not be used to assess any single-laboratory procedures. While a single-laboratory procedure could be approximately applied to the regression design data, this would yield less useful results than by instructing each laboratory to perform the single-laboratory procedure directly. Therefore, it was decided to instruct each laboratory to perform one of the single-laboratory procedures in addition to the regression design laboratory analyses. The Technical Work Group decided that the most appropriate single-laboratory procedure to evaluate in the pilot study is the ACIL L_C/L_Q procedure.

The methodology for confirmation of laboratory analyses initially discussed was to first determine the specific limits for the reviewed procedures, and then instruct labs to spike at each limit (i.e., two-stage confirmation). These additional sample laboratory analyses would then be used to assess whether the calculated limit achieved the required MQO criteria. However, this approach would require a greater amount of time than is available for the study, because the laboratories could not begin analyzing confirmation samples until the limits have been calculated.

Therefore, the possibility of "concurrent confirmation" sample laboratory analyses was discussed. Concurrent sample laboratory analyses would be additional laboratory analyses conducted during the same period as the main regression study laboratory analyses. Because it would not be known what the final limits would be, these additional confirmation laboratory analyses would need to be performed at each of the spike levels. Mean recovery, percent RSD and false negative rates would then be calculated at each spike level using the confirmation data, and models would be fit to estimate these MQO criteria at the determined limits. While the additional laboratory analyses at each concentration would have a greater cost than a two-stage confirmation, the amount of time necessary would be much shorter. In addition, the confirmation sample analyses

from a single laboratory could also be used for assessing the single-laboratory limit determined for that laboratory. Two-stage confirmation could also require a large number of spike laboratory analyses to yield a reliable assessment of the estimated limit.

While the addition of confirmation sample laboratory analyses and the assessment of a single-laboratory procedure will yield more useful information, the additional costs necessitated some changes to the original regression-based pilot design. Based on the experience of the members of the Team, the estimated price per sample was modified for each method; however, the revised costs still exceeded those that could be afforded in the study. The possibility of dropping one of the five methods was first discussed, however it was decided that each of the methods was necessary and should only be dropped as a last resort. Therefore, it was decided that the discussion of possible trade-offs should focus on the numbers of labs, concentrations and replicates to be used in the pilot study.

Because the Regression Group draft pilot design had included seven replicates per spike level to be used to calculate limits, and because it was initially decided that there should be an equal number of replicates used for limit calculation and for limit confirmation, the initial cost estimates were based on fourteen total replicates per spike level. However, no procedure requires more than five replicates per spike level. Therefore, only ten total replicates per spike level (five for calculation, five for confirmation) could be analyzed for each laboratory and still have sufficient data for limit confirmation.

The minimum of nine laboratories included in the Regression Group pilot study design was chosen based on the IDE/IQE requirement for a minimum of six valid laboratories, and the risk of not all laboratories supplying valid data. It was first suggested that only a subset of the laboratories perform the single-laboratory procedure, which would decrease the total costs. However, single-laboratory detection assessments will be performed by laboratories much more frequently than multi-laboratory or interlaboratory detection assessments. Therefore, it is important to include as many assessments of the single-laboratory procedure as possible, and this solution was rejected. Because some volunteer laboratories may also participate in the study, and the possibility of three laboratories supplying invalid data was unlikely, it was decided to drop the number of paid laboratories from 9 to 8.

Additionally, the number of spike levels was dropped from 13 to 12 for the interlaboratory pilot. While a large number of spike levels is needed to cover both the detection and quantitation ranges for both single-laboratory and interlaboratory variability, no procedure requires more than four total spike levels. Therefore, one spike level could be dropped if an appropriate range of spike levels is chosen.

It was originally decided that PCB Aroclors would be analyzed by Method 608 as part of the regression-based pilot study. However, inclusion of PCB Aroclors would potentially require many more sample laboratory analyses, especially when assessing the single-laboratory procedure. Because the MMA PCB study also includes PCB laboratory analyses, it was decided that inclusion of PCBs would not be necessary for the regression-based pilot study. Two Aroclors (Aroclors 1016 and 1260) would still be included in the single-laboratory procedure assessment. Dropping PCB laboratory analyses from the regression pilot study would mean that no data would be available for confirming the calculated single-laboratory limits. Therefore, a smaller number of laboratory analyses for the two Aroclors would need to be performed for single-laboratory limit confirmation. It was decided that this could be accomplished using five replicates at three spike levels for each laboratory. However, the smaller number of replicates would mean that the limit would have to be known prior to starting confirmation laboratory analyses. Therefore, it may not be possible to include the same amount of temporal variability in these confirmation laboratory analyses that was included in the calculation of the single-laboratory limits.

Additional possible changes to the pilot study design were discussed in the event that laboratory bids were larger than expected. In this event, it was decided that one of the methods should be dropped. Method 300.0 was identified as the method that would be most appropriate to drop. As a second option, it was decided that, due to the large amount of existing blank data, Method 200.7 could be dropped from the single-laboratory procedure assessment. Because the LC-MRL, Hubaux-Vos, IDE and IQE procedures can not be performed using only blank data, it was decided that Method 200.7 should not be dropped from the regression-based pilot assessment unless absolutely necessary.

Because it is not known how many laboratories will submit bids, the possibility exists that less than the minimum 8 laboratories will be available to participate in the pilot study for a given method. Because the LC-MRL, Hubaux-Vos and the single lab procedure yield independent limits for each laboratory, no limit exists for the minimum number of laboratories needed to include these procedures in the pilot study. However, the IDE and IQE procedures require a minimum of six laboratories, and therefore could not be determined if fewer than six laboratories participate in the pilot study for a given method.

V. Pilot Confirmation Study – Questions to the Labs Conducting the Pilot

The Team recognized that to make the most from this pilot study, they will need to obtain feedback from the labs doing the pilot testing. Information from the labs could be critical in evaluating the procedures and in assisting the committee with their recommendations. The Technical Work Group agreed with this conclusion.

As part of the EPA solicitation of bids from laboratories, there is a requirement for the labs to respond with a narrative report of their results and process. The narrative report must include:

- Identify and detail any problems associated with the preparation or analysis of specific samples
- Provide comments on performance of the procedures used to calculate detection and quantitation limits in Tasks 2 and 4, specifically addressing the following issues for each procedure in each Task:
  - Clarity of the procedures (i.e., were they clearly written and easy to understand? What areas need clarification?)
  - Required skills for implementing them (i.e., did you find that specialized skills or tools were needed to calculate the limits?)
  - Appropriateness of data (i.e., was the volume and/or type of data required by the procedure appropriate for the limits being determined?)
  - Application of the procedure to other methods (i.e., do you believe the detection and quantitation procedure could be readily applied to other analytical methods that you perform in your lab?)
  - Application to real-world matrices (i.e., Do you believe the procedure would work if it were applied to real-world sample matrices, or that the limits determined in have relevance to real-world sample matrices?)
  - Suggested areas of improvements
- If modifications were requested and pre-approved for use in the study, detail the modifications as they were used in the study
- Include a statement and signature by the Laboratory Manager (or his/her designee) certifying that the analytical and QA/QC data submitted in the data package are accurate, compliant with the terms of the Statement of Work, and complete.

The Technical Work Group agreed that these questions and the narrative should respond to those issues that are important for the committee and the committee efforts at developing recommendations.

VI. Responsibilities, Assignments, Schedule, and Outstanding Decisions

Responsibilities

The Team agreed to the assignment of responsibilities to the many groups involved in the pilot study. These responsibilities are shown in the Table below.

Technical Work Group	• Recommend study design to committee • Promptly respond to questions and problems • Review study results and report • Guide the data analysis to be conducted

EPA OW	• Provide & direct contractor support to implement study • Ensure study goals and deadlines are met • Direct the resolution of non-routine issues that arise during study • Review and approve all study materials • Review and approve study results and report
Study Design Team	• Assist EAD & TWG in finalizing design details • Evaluate proposed lab spiking concentrations and prequalification data • Help develop study spiking scheme (the number of concentrations and exact level of each concentration to be analyzed) • Help develop sample sequence laboratory analysis scheme
Spiking Lab/Vendor	• Prepare and distribute spiked reagent water samples as directed
EPA Contractor (The prime contractor)	• Prepare study materials (study plans, QA plan, lab SOWs/instructions, data reporting formats, checklists, etc.) • Procure and distribute all study materials • Provide lab support services & help recruit volunteers • Draft study spiking scheme and sequence • Day-to-day tracking, communication & resolution of problems w/ labs • Report to EPA and committee on study status, logistics, and technical issues • Review and analyze study results • Draft initial study report
Participant Labs	• Submit recent blanks and calibration data generated with instruments/methods they will use in study • Analyze samples and report data exactly as directed • Immediately notify The prime contractor of delays or problems encountered • Maintain records for 5 years • Respond to follow-up questions

Study Schedule Summary

The Team really focused on concerns about the time availability for the pilot and data analysis and making sure that the committee receives results in a timeframe that allows them the ability to make recommendations. The schedule below is what the Team recommends with the understanding that any significant delays will create a need to revisit the design:

Refine study design – Pilot Design Team	April 27/28,2006
TWG approves study plan	May 16, 2006
Prequalify as part of soliciting bids	June 6 – June 23
Evaluate proposed spiking levels received in pre-qual packages & finalize study spiking scheme	June 23 – July 13
Award to lab bidders providing the best value	Aug 1, 2006
Award spiking vendor	Aug 1, 2006
Prepare and verify conc. of study samples	August 2006
Distribution of study samples and laboratory analysis	Sept. 5 – Oct. 31, 2006
Lab reporting of data	Sept. 29 – Nov. 14, 2006
Data QC review by contractor, and problem resolution	Concurrently with lab reporting through Dec. 20
Data analysis and prepare study report– The prime contractor and the TWG	January – Feb. 15, 2007
Caucus discussions of results and any necessary modifications to procedures	Feb-March, 2007
Recommendations to committee	Late March, 2007
Final committee recommendation	Mid-May, 2007

Outstanding Decisions

The following items need further decisions and will be worked on by the Team and the Technical Work Group in the coming month:

- Understanding different scenarios if the schedule is not maintainable
- Further development of the pilot confirmation of testing
- Data analysis
- Data pooling

ATTACHMENT A
Pilot Design Team Representatives

Zonetta English – Public Utilities
Richard Rediske – Environmental Community
Richard Burrows – Environmental Laboratories
Larry LaFleur – Industry
Dick Reding – EPA
Bob Avery - States

ATTACHMENT B
Draft Pilot Study Purpose and Objectives
Revised December 9, 2005

Background

The Technical Work Group charged two subgroups (Multi-lab and Single-lab) to explore purposes and objectives that a pilot study design should achieve. With input from the two subgroups, the Technical Work Group identified the following purpose and objective statements for presentation and discussion at the December 8-9 committee meeting. Once a pilot study purpose and objective can be defined, the Technical Work Group can begin drafting a definitive study design for the federal advisory committee's consideration.

Purpose

Collect information about various detection and quantitation procedures that will be helpful to the federal advisory committee in its deliberations of detection and quantitation approaches and uses in Clean Water Act programs.

Objectives

Design a study that answers the following pre- and post-study questions:

1. Is the procedure clearly written?
2. Can the data be easily processed in the laboratory?
3. Was the procedure performed correctly?

   [*Note:* the Technical Work Group discussed at length the intent of this question. Members agreed that it was appropriate to have pre- and post-study questions. This question falls into the latter category. It is an attempt to evaluate how the performance of the procedure as written. If there is great variation in the analysis of the performance of the procedure, the Technical Work Group agreed that could be due to variation in interpretation of the written procedure and/or that the procedure was poorly written. The intent is to gauge whether a lab is following and interpreting the procedure correctly, completed through written evaluations from the lab.]

4. How did or will the experimental design influence the outcome of the study?

   Additional clarifying questions from the Multi-lab Subgroup include:

   Type of method (censored, uncensored, etc.)

   a. Works equally well if analyte recoveries are uniformly low, uniformly high, or highly variable
   b. Choice of outlier test (not mandated by procedure?)
   c. Number of different concentrations tested
   d. Number of replicates per concentration tested
   e. Magnitude of concentrations tested

f. Relative relationship between spikes (0.25x, .5x, x, 2x, 4x, etc.)
g. Number of laboratories
h. Number of analysts per study or per laboratory
i. Number and type of instruments per study or per laboratory
j. Sample preparation
k. Number of different days for which analyses are conducted per laboratory
l. Time span over which analyses are conducted per laboratory (week, month, quarter, year)
m. Number of data points per detection or quantitation limit calculation

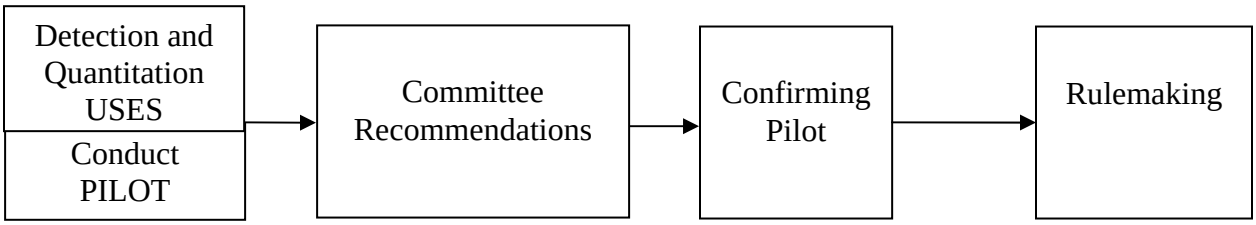
5. Does the procedure achieve its intended purpose?

6. Does the procedure work for all different types of analytical methods?

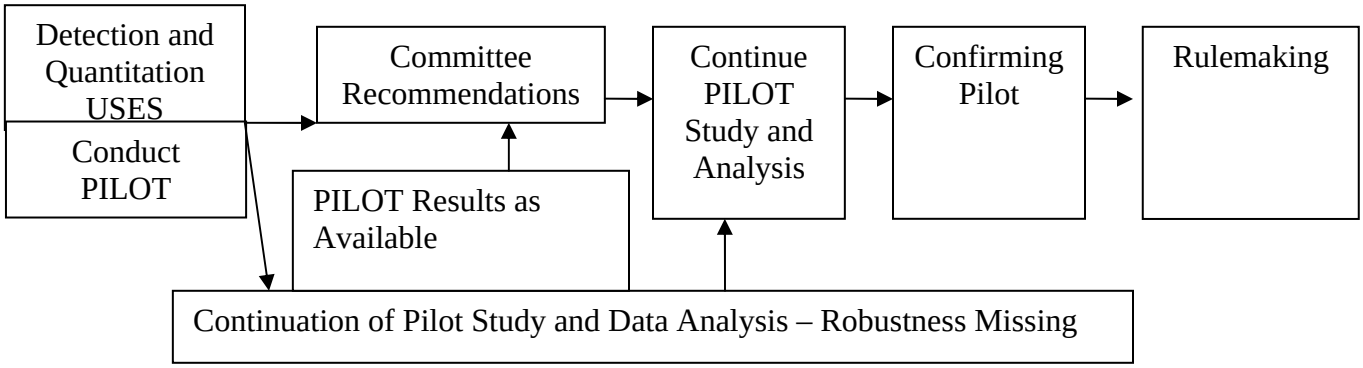
7. Does the procedure work if applied to real world sample matrices? (This may also include a broader question evaluating how the procedure applies to real world matrices.)

ATTACHMENT C
Pilot Study Sequence Options

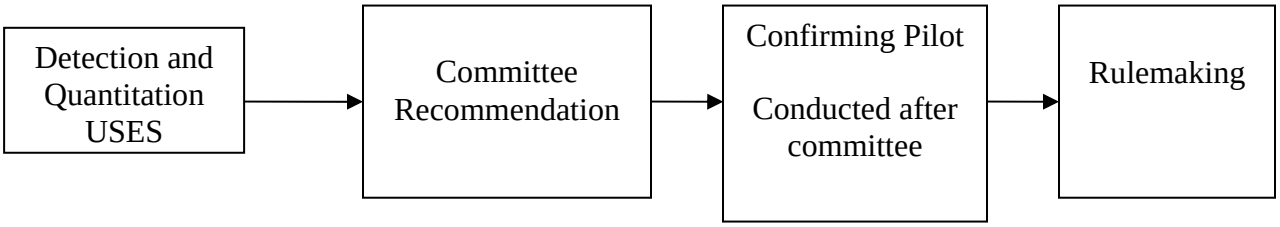
Option A-1



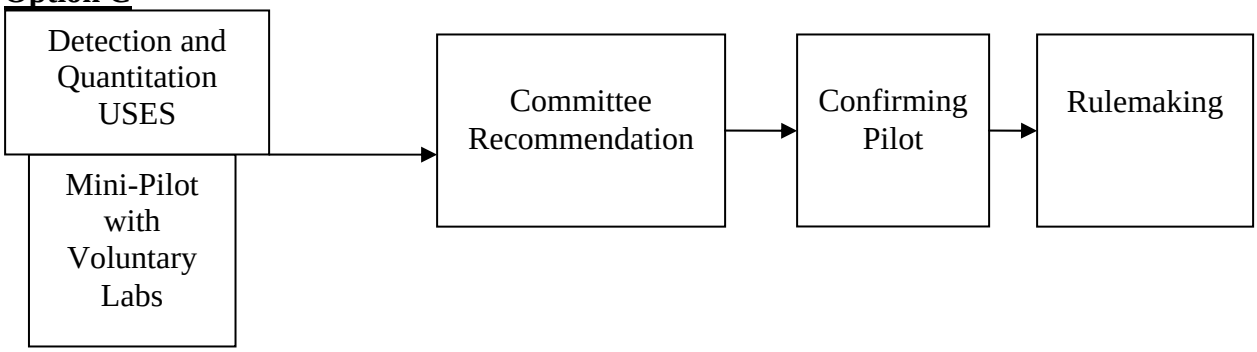
Option A-2



Option B



Option C



ATTACHMENT D
Summary of Details of Pilot Study Design

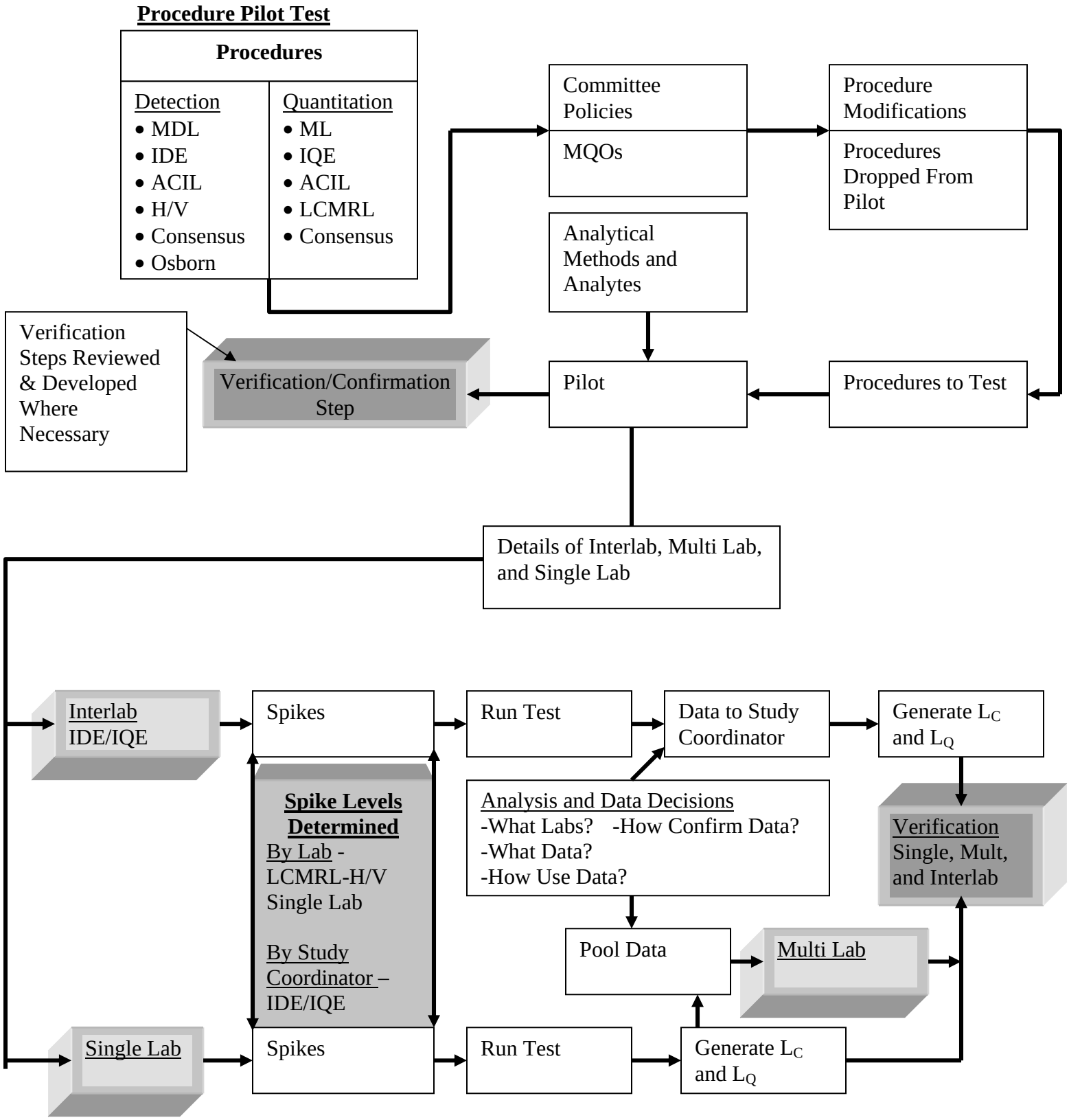
FACA Analytical Database Development – Regression Design

Procedure Type	Instrument	Parameter	Method	Reps/ level	Spike levels	Number of Labs
1	GC/MS	Semi-VOAs	625	10	12	8
2	GC/ECD	Pesticides	608	10	12	8
3	ICP/OES	Metals	200.7	10	12	8
4	IC	Anions	300.0	10	12	8
5	Spec	Cyanides	335.4	10	12	8

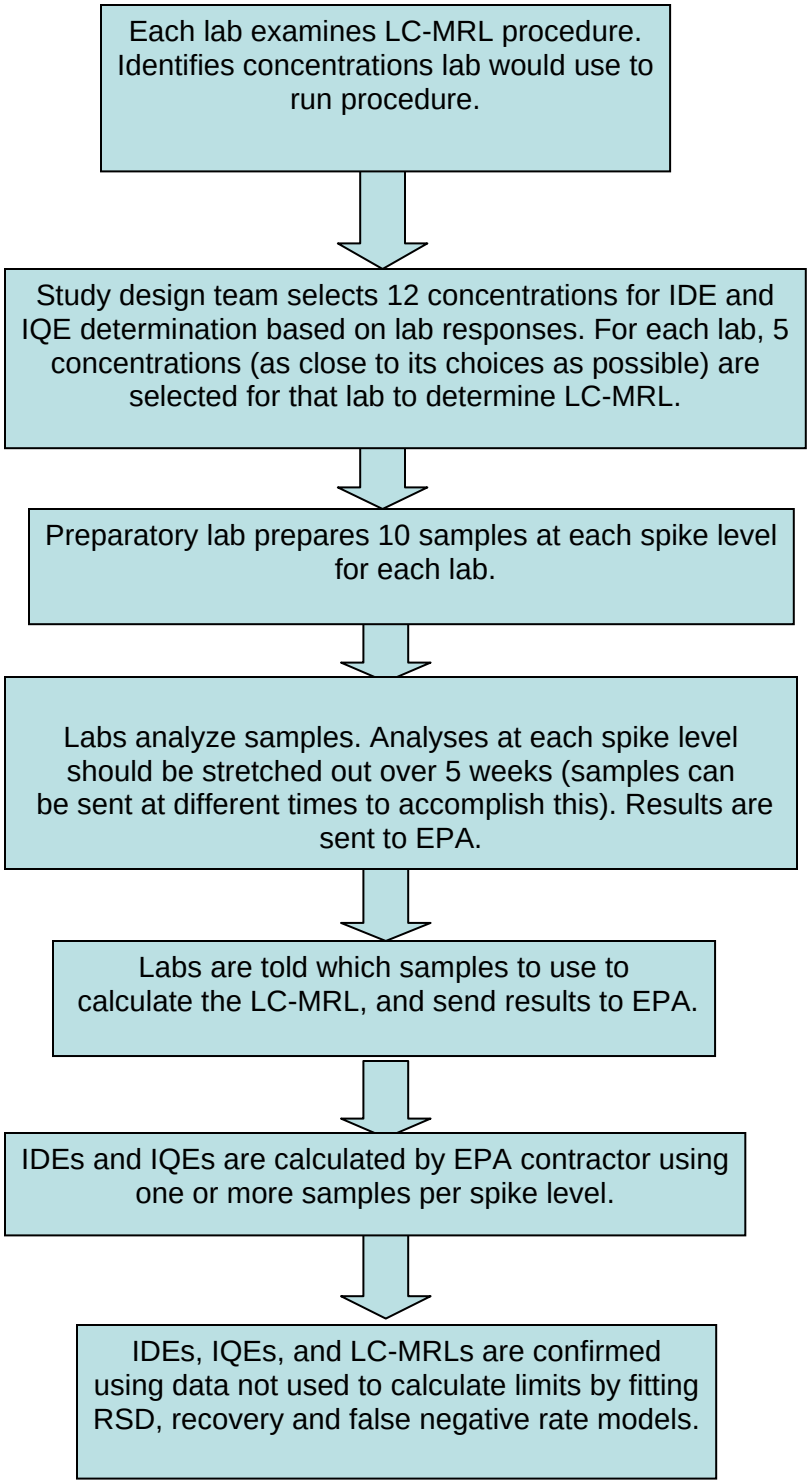
FACA Analytical Database Development - Single Lab Design

1	GC/MS	Semi-VOAs	625	27	3	8
2	GC/ECD	Pest/PCB	608	27	3	8
3	ICP/OES	Metals	200.7	27	3	8
4	IC	Anions	300.0	27	2	8
5	Spec-Auto	Cyanides	335.4	27	1	8
	Single lab PCB L_Q Confirmation		M608-Aroclors	5	3	8

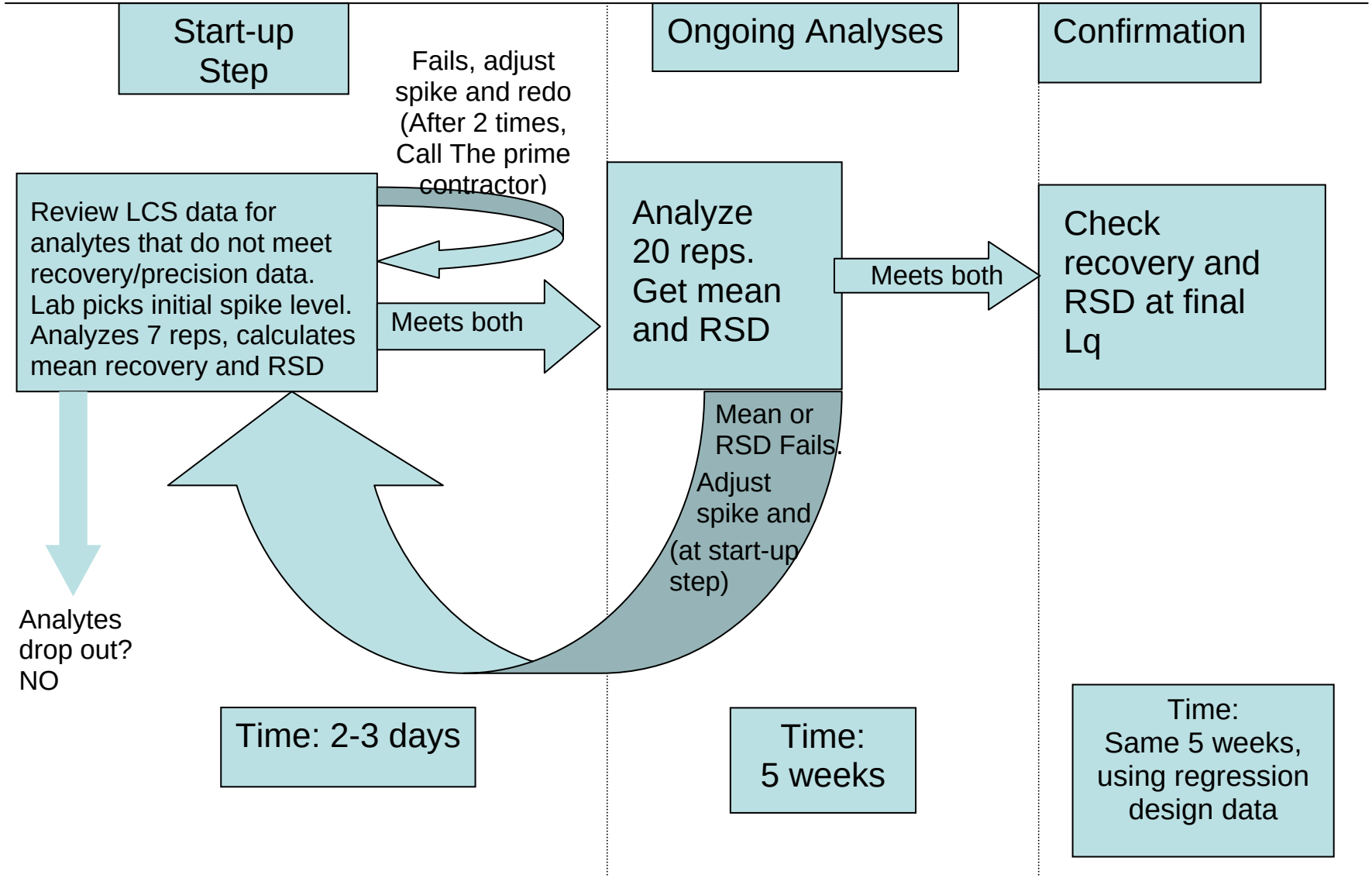
ATTACHMENT E
Pilot Design Schematic
05/01/06



Attachment F
Process for Regression-Based Pilot Study



ATTACHMENT G
Process for Pilot Study Evaluation of Single-lab Procedure



ATTACHMENT H
Target Analytes for Pilot Study Design

Method 608
Aldrin
Alpha-BHC
Beta-BHC
Delta-BHC
Gamma-BHC
Alpha-Chlordane
Gamma-Chlordane
4,4'-DDD
4,4'-DDE
4,4'-DDT .
Dieldrin
Endosulfan I
Endosulfan II
Endosulfan sulfate
Endrin
Endrin aldehyde
Heptachlor
Heptachlor epoxide
PCB-1016
PCB-1260

Method 625
Acenaphthene .....
Acenaphthylene .....
Anthracene .....
Benzo(a)anthracene .....
Benzo(b)fluoranthene .....
Benzo(k)fluoranthene .....
Benzo(a)pyrene .....
Benzo(ghi)perylene .....
Benzyl butyl phthalate .....
Bis(2-chloroethyl)ether.....
Bis(2-chloroethoxy)methane.....
Bis(2-ethylhexyl)phthalate .....
Bis(2-chloroisopropyl)ether .....
4-Bromophenyl phenyl ether .....
2-Chloronaphthalele .....
4-Chlorophenyl phenyl ether.....
Chrysene .....

Method 625
Dibenzo(a,h)anthracene .....
Di-n-butylphthalate .....
1,3-Dichlorobenzene .....
1,2-Dichlorobenzene .....
1,4-Dichlorobenzene .....
3,3'-Dichlorobenzidine .....
Dieldrin.....
Diethyl phthalate.....
Dimethyl phthalate .....
2,4-Dinitrotoluene .....
2,6-Dinitrotoluene .....
Di-n-octylphthalate .....
Fluoranthene .....
Fluorene .....
Hexachlorobenzene .....
Hexachlorobutadiene.....
Hexachloroethane .....
Indeno(1,2,3-cd)pyrene .....
Isophorone .....
Naphthalene .....
Nitrobenzene.....
N-Nitrosodi-n-propylamine.....
Phenanthrene .....
Pyrene.....
1,2,4-Trichlorobenzene.....
4-Chloro-3-methylphenol.....
2-Chlorophenol .....
2,4-Dichlorophenol .....
2,4-Dimethylphenol .....
2,4-Dinitrophenol .....
2-Methyl-4,6-dinitrophenol .....
2-Nitrophenol .....
4-Nitrophenol .....
Pentachlorophenol.....
Phenol.....
2,4,6-Trichlorophenol.....

Method 200.7

Aluminum
Antimony
Arsenic
Barium

Method 200.7
Beryllium

Cadmium
Calcium
Chromium
Cobalt
Copper
Iron
Lead

Magnesium
Manganese

Molybdenum
Nickel
Phosphorus
Potassium
Selenium

Silver
Sodium

Thallium
Tin
Titanium
Vanadium
Zinc

Method 300.0
Bromide
Chloride
Fluoride
Nitrate-N
Nitrite-N

Otho-Phosphate-P

Sulfate